

LIMITING SPECTRAL DISTRIBUTION OF A NEW RANDOM MATRIX MODEL WITH DEPENDENCE ACROSS ROWS AND COLUMNS

OLIVER PFAFFEL AND ECKHARD SCHLEMM

ABSTRACT. We introduce a random matrix model where the entries are dependent across both rows and columns. More precisely, we investigate matrices of the form $\mathbf{X} = (X_{(i-1)n+j})_{it} \in \mathbb{R}^{p \times n}$ derived from a linear process $X_t = \sum_j c_j Z_{t-j}$, where the $\{Z_t\}$ are independent random variables with bounded fourth moments. We show that, when both p and n tend to infinity such that the ratio p/n converges to a finite positive limit y , the empirical spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^T$ converges almost surely to a deterministic measure. This limiting measure, which depends on y and the spectral density of the linear process X_t , is characterized by an integral equation for its Stieltjes transform. The matrix $p^{-1}\mathbf{X}\mathbf{X}^T$ can be interpreted as an approximation to the sample covariance matrix of a high-dimensional process whose components are independent copies of X_t .

1. INTRODUCTION

Random matrix theory studies the properties of large random matrices $A = (A_{i,j})_{ij} \in \mathbb{K}^{p \times n}$, for some field \mathbb{K} . In this article, the entries A_{ij} are real random variables unless otherwise specified. Commonly, the focus is on asymptotic properties of such matrices as their dimensions tend to infinity. One particularly interesting object of study is the asymptotic distribution of their singular values. Since the squared singular values of A are the eigenvalues of AA^T , this is often done by investigating the eigenvalues of AA^T , which is called a sample covariance matrix. The spectral characteristics of a $p \times p$ matrix S are conveniently studied via its empirical spectral distribution, which is defined as $F^S = p^{-1} \sum_{i=1}^p \delta_{\lambda_i}$; here, $\{\lambda_1, \dots, \lambda_p\}$ are the eigenvalues of S , and δ_x denotes the Dirac measure located at x . For some set $B \subset \mathbb{R}$, the figure $F^S(B)$ is the number of eigenvalues of S that lie in B . The measure F^S is considered a random element of the space of probability distributions equipped with the weak topology, and we are interested in its limit as both n and p tend to infinity such that the ratio p/n converges to a finite positive limit y .

The first result of this kind can be found in the remarkable paper of Marchenko and Pastur [14]. They showed that $F^{p^{-1}AA^T}$ converges to a non-random limiting spectral distribution $\hat{F}^{p^{-1}AA^T}$ if all A_{ij} are independent, identically distributed, centred random variables with finite fourth moment. Interestingly, the Lebesgue density of $\hat{F}^{p^{-1}AA^T}$ is given by an explicit formula which only involves the ratio y and the common variance of A_{ij} and is therefore universal with respect to the distribution of the entries of A . Subsequently [22, 24], the same result was obtained under the weaker moment condition that the entries A_{ij} have finite variance. The requirement that the entries of A be identically distributed has later been relaxed to a Lindeberg-type condition, cf. Eq. (3). For more details and a comprehensive treatment of random matrix theory we refer the reader to the text books Anderson et al. [1], Bai and Silverstein [5], Mehta [16].

Key words and phrases. eigenvalue distribution; limiting spectral distribution; linear process; random matrix theory; random matrix with dependent entries; sample covariance matrix.

Both authors gratefully acknowledge financial support from Technische Universität München - Institute for Advanced Study funded by the German Excellence Initiative and from the International Graduate School of Science and Engineering.

Recent research has focused on the question to what extent the assumption of independence of the entries of A can be relaxed without compromising the validity of the Marchenko–Pastur law. In Aubrun [3] it was shown that for random matrices A whose rows are independent \mathbb{R}^n -valued random variables uniformly distributed on the unit ball of $l_q(\mathbb{R}^n)$, $q > 1$, the empirical spectral distribution $F^{p^{-1}AA^\top}$ still converges to the same law as in the i. i. d. case. The Marchenko–Pastur law is, however, not stable with respect to more substantial deviations from the independence assumptions.

A very useful tool to characterize the limiting spectral distribution in random matrix models with dependent entries is the Stieltjes transform which, for some measure μ , is defined as the map $s_\mu : \mathbb{C}^+ \rightarrow \mathbb{C}^+$, $s_\mu(z) = \int_{\mathbb{R}} (t - z)^{-1} \mu(dt)$. A particular, very successful random matrix model exhibiting dependence within the rows was investigated already by Marchenko and Pastur [14] and later in greater generality by Pan [17], Silverstein and Bai [20]: they modelled dependent data as a linear transformation of independent random variables which led to the study of the eigenvalues of random matrices of the form AHA^\top , where the entries of A are independent, and H is a positive semidefinite population covariance matrix whose spectral distribution converges to a non-random limit \hat{F}^H . They found that the Stieltjes transform of the limiting spectral distribution of $p^{-1}AHA^\top$ can be characterized as the solution to an integral equation involving only \hat{F}^H and the ratio $y = \lim p/n$. Another approach, suggested in Bai and Zhou [4] and further pursued in Pfaffel and Schlemm [18], is to model the rows of A independently as stationary linear processes with independent innovations. This structure is interesting because the class of linear processes includes many practically relevant time series models, such as (fractionally integrated) ARMA processes, as special cases. The main result of Pfaffel and Schlemm [18] shows that for this model the limiting spectral distribution depends only on y and the second-order properties of the underlying linear process.

All results for independent rows with dependent row entries also hold with minor modifications for the case where A has independent columns with dependent column entries. This is due to the fact that the matrices AA^\top and $A^\top A$ have the same non-zero eigenvalues.

In contrast, there are only very few results dealing with random matrix models where the entries are dependent across both rows and columns. The case where A is given as the result of a two-dimensional linear filter applied to an array of independent complex Gaussian random variables is considered in Hachem et al. [10]. They use the fact that A can be transformed to a random matrix with uncorrelated, non-identically distributed entries. Because of the assumption of Gaussianity the entries are in fact independent, and so an earlier result by the same authors [11] can be used to obtain the asymptotic distribution of the eigenvalues of $p^{-1}AA^*$. In the context of operator-valued free probability theory, Rashidi Far et al. [19] succeeded in characterizing the limiting spectral distribution of block Wishart matrices through a quadratic matrix equation for the corresponding operator-valued Stieltjes transform.

A parallel line of research focuses on the spectral statistics of large symmetric or Hermitian square matrices with dependent entries, thus extending Wigner’s [23] seminal result for the i. i. d. case. Models studied in this context include random Toeplitz, Hankel and circulant matrices [6, 8, 15, and references therein] as well as approaches allowing for a more general dependence structure [2, 12].

In Pfaffel and Schlemm [18], the authors considered sample covariance matrices of high-dimensional stochastic processes, the components of which are modelled by independent infinite-order moving average processes with identical second-order characteristics. In practice, it is often not possible to observe all components of such a high-dimensional process, and the sample covariance matrix can then not be computed. To solve this problem when only one component is observed, it seems reasonable to partition one long observation record of that observed component of length pn into p segments of length n , and to treat the different segments as if they were records of the unobserved components.

We show that this approach is valid and leads to the correct asymptotic eigenvalue distribution of the sample covariance matrix if the components of the underlying process are modelled as independent moving averages.

We are thus led to investigate a model of random matrices \mathbf{X} whose entries are dependent across both rows and columns, and which is not covered by the results mentioned above. The entries of the random matrix under consideration are defined in terms of a single linear stochastic process, see Section 2 for a precise definition. Without assuming Gaussianity we prove almost sure convergence of the empirical spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^\top$ to a deterministic limiting measure and characterize the latter via an integral equation for its Stieltjes transform, which only depends on the asymptotic aspect ratio of the matrix and the second-order properties of the underlying linear process. Our result extends the class of random matrix models for which the limiting spectral distribution can be identified explicitly by a new, theoretically appealing model. It thus contributes to laying the ground for further research into more general random matrix models with dependent, non-identically distributed entries. **Outline.** In Section 2 we give a precise definition of the random matrix model we investigate and state the main result about its limiting spectral distribution. The proof of the main theorem as well as some auxiliary results are presented in Section 3. Finally, in Section 4, we indicate how our result could be obtained in an alternative way from a similar random matrix model with independent rows.

Notation. We use \mathbb{E} and $\mathbb{V}\text{ar}$ to denote expected value and variance. Where convenient, we also write $\mu_{1,X}$ and $\mu_{2,X}$ for the first and second moment, respectively, of a random variable X . The symbol $\mathbf{1}_m$, m a natural number, stands for the $m \times m$ identity matrix. For the trace of a matrix S we write $\text{tr } S$. For sequences of matrices $(S_n)_n$ we will suppress the dependence on n where this does not cause ambiguity; the sequence of associated spectral distributions is denoted by F^S , and for their weak limit, provided it exists, we write \hat{F}^S . It will also be convenient to use asymptotic notation: for two sequences of real numbers $(a_n)_n, (b_n)_n$ we write $a_n = O(b_n)$ to indicate that there exists a constant C which is independent of n , such that $a_n \leq Cb_n$ for all n . We denote by \mathbb{Z} the set of integers and by \mathbb{N}, \mathbb{R} , and \mathbb{C} the sets of natural, real, and complex numbers, respectively. $\Im z$ stands for the imaginary part of a complex number z , and \mathbb{C}^+ is defined as $\{z \in \mathbb{C} : \Im z > 0\}$. The indicator of an expression \mathcal{E} is denoted by $I_{\{\mathcal{E}\}}$ and defined to be one if \mathcal{E} is true and zero otherwise.

2. A NEW RANDOM MATRIX MODEL

For a sequence $(Z_t)_{t \in \mathbb{Z}}$ of independent real random variables and real coefficients $(c_j)_{j \in \mathbb{N} \cup \{0\}}$, the linear process $(X_t)_{t \in \mathbb{Z}}$ and the $p \times n$ matrix \mathbf{X} are defined by $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$ and

$$\mathbf{X} = (\mathbf{X}_{i,t})_{i,t} = (X_{(i-1)n+t})_{i,t} = \begin{pmatrix} X_1 & \dots & X_n \\ X_{n+1} & \dots & X_{2n} \\ \vdots & & \vdots \\ X_{(p-1)n+1} & \dots & X_{pn} \end{pmatrix} \in \mathbb{R}^{p \times n}. \quad (1)$$

The interesting feature about this matrix \mathbf{X} is that its entries are dependent across both rows and columns. In contrast to models considered in [4, 11, 18], not all entries far away from each other are asymptotically independent, e. g., the correlation between the entries $\mathbf{X}_{i,n}$ and $\mathbf{X}_{i+1,1}$, $i = 1, \dots, p-1$, does not depend on n . We will investigate the asymptotic distribution of the eigenvalues of $p^{-1}\mathbf{X}\mathbf{X}^\top$ as both p and n tend to infinity such that their ratio p/n converges to a finite, positive limit y . We assume that the sequence $(Z_t)_t$ satisfies

$$\mathbb{E}Z_t = 0, \quad \mathbb{E}Z_t^2 = 1, \quad \text{and} \quad \sigma_4 := \sup_t \mathbb{E}Z_t^4 < \infty, \quad (2)$$

and that the following Lindeberg-type condition is satisfied: for each $\epsilon > 0$,

$$\frac{1}{pn} \sum_{t=1}^{pn} \mathbb{E} \left(Z_t^2 I_{\{Z_t^2 \geq \epsilon n\}} \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (3)$$

Condition (3) is satisfied if all $\{Z_t\}$ are identically distributed, but that is not necessary. As it turns out, the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^\top$ depends only on y and the second-order structure of the underlying linear process X_t , which we now recall: its auto-covariance function $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ is defined by $\gamma(h) = \mathbb{E}X_0X_h = \sum_{j=0}^{\infty} c_j c_{j+|h|}$; its spectral density $f : [0, 2\pi] \rightarrow \mathbb{R}$ is the Fourier transform of γ , namely $f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ih\omega}$. The following is the main result of the paper.

Theorem 1. *Let $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with continuously differentiable spectral density f , and let the matrix $\mathbf{X} \in \mathbb{R}^{p \times n}$ be given by Eq. (1). Assume that*

- i) *the sequence $(Z_t)_t$ satisfies conditions (2) and (3),*
- ii) *there exist positive constants C, δ such that $|c_j| \leq C(j+1)^{-1-\delta}$, for all $j \in \mathbb{N} \cup \{0\}$.*

Then, as n and p tend to infinity such that the ratio p/n converges to a finite positive limit y , the empirical spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^\top$ converges almost surely to a non-random probability distribution \hat{F} with bounded support. The Stieltjes transform $z \mapsto s_{\hat{F}}(z)$ of \hat{F} is the unique mapping $\mathbb{C}^+ \rightarrow \mathbb{C}^+$ satisfying

$$\frac{1}{s_{\hat{F}}(z)} = -z + y \int_0^{2\pi} \frac{f(\omega)}{1 + f(\omega)s_{\hat{F}}(z)} d\omega. \quad (4)$$

Remark 1. The assumption that the coefficients $(c_j)_j$ decay at least polynomially is not very restrictive; it allows, e. g., for X_t to be an ARMA or fractionally integrated ARMA process, which exhibits long-range dependence [9, 13]. In the latter case the entries of the matrix \mathbf{X} are long-range dependent as well.

Remark 2. It is possible to generalize the proof of Theorem 1 so that the result also holds for non-causal processes, where $X_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j}$. The required changes are merely notational, the only difference in the result is that the auto-covariance function is then given by $\sum_{j=-\infty}^{\infty} c_j c_{j+|h|}$.

The distribution \hat{F} can be obtained from $s_{\hat{F}}$ via the Perron–Frobenius inversion formula [5, Theorem B.8], which states that for all continuity point $0 < a < b$ of \hat{F} , it holds that $\hat{F}([a, b]) = \lim_{\epsilon \rightarrow 0^+} \int_a^b \Im s_{\hat{F}}(x + \epsilon i) dx$. In general, the analytic determination of this distribution is not feasible. It is, however, easy to check that for the special case of independent entries one recovers the classical Marchenko–Pastur law.

3. PROOF OF THEOREM 1

The strategy in the proof of Theorem 1 is to show that the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^\top$ is stable under modifications of \mathbf{X} which reduce the sample covariance matrix to the form $p^{-1}\mathbf{Z}\mathbf{H}\mathbf{Z}^\top$, for a matrix \mathbf{Z} with i. i. d. entries, and some positive definite \mathbf{H} . To this end we will repeatedly use the following lemma which presents sufficient conditions for the limiting spectral distributions of two sequences of matrices to be equal.

Lemma 2 (Trace criterion). *Let $A_{1,n}, A_{2,n}$ be sequences of $p \times n$ matrices, where $p = p_n$ depends on n such that $p_n \rightarrow \infty$ as $n \rightarrow \infty$. Assume that the spectral distribution $F^{p^{-1}A_{1,n}A_{1,n}^\top}$ converges almost surely to a deterministic limit $\hat{F}^{p^{-1}A_{1,n}A_{1,n}^\top}$ as n tends to infinity. If there exists a positive number ϵ such that*

- i) $p^{-4} \mathbb{E} \left[\text{tr} (A_{1,n} - A_{2,n}) (A_{1,n} - A_{2,n})^\top \right]^2 = O(n^{-1-\epsilon}),$
- ii) $p^{-2} \mathbb{E} \text{tr} A_{i,n} A_{i,n}^\top = O(1), i = 1, 2, \text{ and}$
- iii) $p^{-4} \mathbb{V} \text{ar} \text{tr} A_{i,n} A_{i,n}^\top = O(n^{-1-\epsilon}), i = 1, 2,$

then the spectral distribution of $p^{-1} A_{2,n} A_{2,n}^\top$ is convergent almost surely with the same limit $\hat{F}^{p^{-1} A_{1,n} A_{1,n}^\top}$.

Proof. The claim is a direct consequence of Chebyshev's inequality, the first Borel–Cantelli lemma, and Bai and Silverstein [5, Corollary A.42] \square

With the constants C and δ from assumption ii) of Theorem 1 we define $\bar{c}_j := C(j+1)^{-1-\delta}$, such that $|c_j| \leq \bar{c}_j$ for all j . Without further reference we will repeatedly use the fact that $j \mapsto \bar{c}_j$ is monotone, that $\sum_{j=1}^{\infty} \bar{c}_j^\alpha$ is finite for every $\alpha \geq 1$, and that $\sum_{j=n}^{\infty} \bar{c}_j^\alpha$ is of order $O(n^{1-\alpha(1+\delta)})$. Since it is difficult to deal with infinite-order moving averages processes directly, it is convenient to truncate the entries of the matrix \mathbf{X} by defining $\tilde{X}_t = \sum_{j=0}^n c_j Z_{t-j}$ and $\tilde{\mathbf{X}} = (\tilde{X}_{(i-1)n+t})_{it}$; this is different from the usual truncation of the support of the entries of a random matrix.

Proposition 3 (Truncation). *If the empirical spectral distribution of $p^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^\top$ converges to a limit, then the empirical spectral distribution of $p^{-1} \mathbf{X} \mathbf{X}^\top$ converges to the same limit.*

Proof. The proof proceeds in two steps in which we verify conditions i) to iii) of Lemma 2.

Step 1. The definitions of \mathbf{X} and $\tilde{\mathbf{X}}$ imply that

$$\Delta_{\mathbf{X}, \tilde{\mathbf{X}}} := \frac{1}{p^2} \text{tr} (\mathbf{X} - \tilde{\mathbf{X}}) (\mathbf{X} - \tilde{\mathbf{X}})^\top = \frac{1}{p^2} \sum_{i=1}^p \sum_{t=1}^n [\mathbf{X}_{it} - \tilde{\mathbf{X}}_{it}]^2 = \frac{1}{p^2} \sum_{i,t=1}^{p,n} \sum_{k,k'=n+1}^{\infty} Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} c_k c_{k'}.$$

We shall show that the second moment of $\Delta_{\mathbf{X}, \tilde{\mathbf{X}}}$ is of order at most $n^{-2-2\delta}$. Since

$$\sum_{\substack{k,k' \\ m,m'=n+1}}^{\infty} \mathbb{E} |Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'}| |c_k| |c_{k'}| |c_m| |c_{m'}| \leq \sigma_4 \left[\sum_{k=0}^{\infty} |c_k| \right]^4 < \infty, \quad (5)$$

we can apply Fubini's theorem to interchange expectation and summation in the computation of

$$\mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}} := \mathbb{E} \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^2 = \frac{1}{p^4} \sum_{\substack{i,i'=1 \\ t,t'=1}}^{p,n} \sum_{\substack{k,k' \\ m,m'=n+1}}^{\infty} \mathbb{E} [Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'}] c_k c_{k'} c_m c_{m'}. \quad (6)$$

Since the $\{Z_t\}$ are independent, the expectation in that sum is non-zero only if all four Z are the same or else one can match the indices in two pairs. In the latter case we distinguish three cases according to which factor the first Z is paired with. This leads to the additive decomposition

$$\mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}} = \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}^{\sqcap \sqcap \sqcap \sqcap} + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}^{\sqcap \sqcap \sqcap \sqcap} + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}^{\sqcap \sqcap \sqcap \sqcap} + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}^{\sqcap \sqcap \sqcap \sqcap}, \quad (7)$$

where the ideograms indicate which of the four factors are equal. For the contribution from all four Z being equal it holds that $k = k'$, $m = m'$, and $(i-1)n+t-k = (i'-1)n+t'-m$, so that

$$\mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}^{\sqcap \sqcap \sqcap \sqcap} = \frac{\sigma_4}{p^4} \sum_{i,i'}^p \sum_{t,t'=1}^n \sum_{m=\max\{n+1, n+1-(i-i')n-(t-t')\}}^{\infty} c_{(i-i')n+(t-t')+m}^2 c_m^2.$$

If we introduce the new summation variables $\delta_i := i - i'$ and $\delta_t := t - t'$, we obtain

$$\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod} = \frac{\sigma_4}{p^4} \sum_{\delta_i=1-p}^{p-1} \underbrace{(p - |\delta_i|)}_{\leq p} \sum_{\delta_t=1-n}^{n-1} \underbrace{(n - |\delta_t|)}_{\leq n} \sum_{m=\max\{n+1, n+1-\delta_i n - \delta_t\}}^{\infty} c_{m+\delta_i n + \delta_t}^2 c_m^2.$$

If δ_i is positive, then $\delta_i n + \delta_t$ is positive as well; the fact that $|c_j|$ is bounded by \bar{c}_j and the monotonicity of $j \mapsto \bar{c}_j$ imply that $c_{m+\delta_i n + \delta_t}^2 \leq \bar{c}_{(\delta_i-1)n} \bar{c}_{\delta_t n}$ so that the contribution from $\delta_i \geq 1$ can be estimated as

$$\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod,+} \leq \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \underbrace{\sum_{\delta_i=1}^{p-1} \bar{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})} \underbrace{\sum_{\delta_t=1}^{2n-1} \bar{c}_{\delta_t}}_{=O(1)} \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^2}_{=O(n^{-1-2\delta})} = O(n^{-4-3\delta}).$$

An analogous argument shows that the contribution from $\delta_i \leq -1$, denoted by $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod,-}$, is of the same order of magnitude. The contribution to $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$ from $\delta_i = 0$ is given by

$$\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod,\emptyset} = \frac{\sigma_4 n}{p^3} \sum_{\delta_t=1-n}^{n-1} \sum_{m=\max\{n+1, n+1-\delta_t\}}^{\infty} c_m^2 c_{m+\delta_t}^2 \leq \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \left[2 \underbrace{\sum_{\delta_t=1}^{n-1} \bar{c}_{\delta_t}^2}_{=O(1)} \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^2}_{=O(n^{-1-2\delta})} + \underbrace{\sum_{m=n+1}^{\infty} \bar{c}_m^4}_{=O(n^{-3-4\delta})} \right] = O(n^{-3-2\delta}).$$

By combining the last two displays, it follows that $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$ is of order $O(n^{-3-2\delta})$. The second term in Eq. (7) corresponds to $k = k'$, $m = m'$, and $(i-1)n + t - k \neq (i'-1)n + t' - m$. The restriction that not all four factors be equal is taken into account by subtracting $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$; consequently,

$$\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod} = \frac{1}{p^4} \underbrace{\sum_{i,i'=1}^p \sum_{t,t'=1}^n}_{=O(1)} \underbrace{\sum_{k,m=n+1}^{\infty} c_k^2 c_m^2}_{=O(n^{-2-4\delta})} - \mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod} = O(n^{-2-4\delta}).$$

It remains to analyse $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$ which, by symmetry, is equal to $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$. If the first factor is paired with the third, the condition for non-vanishment becomes $k = m + (i - i')n + t - t'$, $k' = m' + (i - i')n + t - t'$, and $m \neq m'$. Again introducing the new summation variables $\delta_i := i - i'$ and $\delta_t := t - t'$, we obtain that

$$\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod} = \frac{1}{p^4} \sum_{\delta_i=1-p}^{p-1} \underbrace{(p - |\delta_i|)}_{\leq p} \sum_{\delta_t=1-n}^{n-1} \underbrace{(n - |\delta_t|)}_{\leq n} \sum_{m,m'=\max\{n+1, n+1-\delta_i n - \delta_t\}}^{\infty} c_m c_{m'} c_{m+\delta_i n + \delta_t} c_{m'+\delta_i n + \delta_t} - \mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}.$$

As in the analysis of $\mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod}$ we obtain the contribution from $\delta_i \neq 0$ as

$$\left| \mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod,+} \right| = \left| \mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod,-} \right| \leq \underbrace{\frac{n}{p^3}}_{=O(n^{-2})} \underbrace{\sum_{\delta_i=1}^{p-1} \bar{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})} \underbrace{\sum_{\delta_t=1}^{2n-1} \bar{c}_{\delta_t}}_{=O(1)} \underbrace{\sum_{m,m'=n+1}^{\infty} \bar{c}_m \bar{c}_{m'}}_{=O(n^{-2\delta})} + \mu_{2,\Delta_{\mathbf{x},\bar{\mathbf{x}}}}^{\prod\prod} = O(n^{-3-2\delta}). \quad (8)$$

Finally, for the contribution from $\delta_i = 0$ one finds that

$$\begin{aligned} \left| \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square, \emptyset}} \right| &\leq \frac{n}{p^3} \sum_{\delta_i=1-n}^{n-1} \sum_{m, m'=\max\{n+1, n+1-\delta_i\}}^{\infty} |c_m c_{m'} c_{m+\delta_i} c_{m'+\delta_i}| + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square}} \\ &\leq \underbrace{\frac{n}{p^3}}_{=O(n^{-2})} \left[\underbrace{2 \sum_{\delta_i=1}^{n-1} \bar{c}_{\delta_i}^2}_{=O(1)} \underbrace{\sum_{m, n'=n+1}^{\infty} \bar{c}_m \bar{c}_{m'}}_{=O(n^{-2\delta})} + \underbrace{\sum_{m, n'=n+1}^{\infty} \bar{c}_m^2 \bar{c}_{m'}^2}_{=O(n^{-2-4\delta})} \right] + \mu_{2, \Delta}^{\square} = O(n^{-2-2\delta}). \end{aligned} \quad (9)$$

The last two displays (8) and (9) imply that $\mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square}} = \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square, -}} + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square, \emptyset}} + \mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}^{\square, +}} = O(n^{-2-2\delta})$. Thus, $\mu_{2, \Delta_{\mathbf{X}, \tilde{\mathbf{X}}}}$ is of order $O(n^{-2-2\delta})$, as claimed.

Step 2. Next we verify assumptions ii) and iii) of Lemma 2, which means that we show that both $\Sigma_{\mathbf{X}} := p^{-2} \text{tr } \mathbf{X} \mathbf{X}^T$ and $\Sigma_{\tilde{\mathbf{X}}} := p^{-2} \text{tr } \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$ have bounded first moments and variances of order $n^{-1-\epsilon}$, for some $\epsilon > 0$; in fact, ϵ will turn out to be one. For $\Sigma_{\mathbf{X}}$ we obtain

$$\mu_{1, \Sigma_{\mathbf{X}}} := \mathbb{E} \Sigma_{\mathbf{X}} = \frac{1}{p^2} \sum_{i=1}^p \sum_{t=1}^n \sum_{k, k'=0}^{\infty} \mathbb{E} [Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'}] c_k c_{k'} = \frac{n}{p} \sum_{k=0}^{\infty} c_k^2,$$

where the change of the order of expectation and summation is valid by Fubini's theorem. Using Eq. (5) and Fubini's theorem, the second moment of $\Sigma_{\mathbf{X}}$ becomes

$$\mu_{2, \Sigma_{\mathbf{X}}} := \mathbb{E} \Sigma_{\mathbf{X}}^2 = \frac{1}{p^4} \sum_{i, i'=1}^p \sum_{t, t'=1}^n \sum_{k, k'=0}^{\infty} \mathbb{E} [Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'}] c_k c_{k'} c_m c_{m'}.$$

This sum coincides with the expression analysed in Eq. (6), except that here the k, k', m, m' sums start at zero, and not at $n+1$. A straightforward adaptation of the arguments there show that $\mu_{2, \Sigma_{\mathbf{X}}}$ equals $n^2 p^{-2} \left(\sum_{k=0}^{\infty} c_k^2 \right)^2 + O(n^{-2})$, and, consequently, that $\text{Var } \Sigma_{\mathbf{X}} = \mu_{2, \Sigma_{\mathbf{X}}} - (\mu_{1, \Sigma_{\mathbf{X}}})^2 = O(n^{-2})$. Analogous computations show that $\mathbb{E} \Sigma_{\tilde{\mathbf{X}}}$ is bounded, and that $\text{Var } \Sigma_{\tilde{\mathbf{X}}} = O(n^{-2})$. Thus, conditions ii) and iii) of Lemma 2 are verified, and the proof of the proposition is complete. \square

Because of Proposition 3 the problem of determining the limiting spectral distribution of the sample covariance matrix $p^{-1} \mathbf{X} \mathbf{X}^T$ has been reduced to computing the limiting spectral distribution of $p^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$, where now, for fixed n , the matrix $\tilde{\mathbf{X}}$ depends on only finitely many of the noise variables Z_t . The fact that the entries of $\tilde{\mathbf{X}}$ are finite-order moving average processes and therefore linearly dependent on the Z_t allows for $\tilde{\mathbf{X}}$ to be written as a linear transformation of the i.i.d. matrix $\mathbf{Z} := (Z_{(i-2)n+t})_{i=1, \dots, p+1, t=1, \dots, n}$. We emphasize that \mathbf{Z} , in contrast to \mathbf{X} and $\tilde{\mathbf{X}}$, is a $(p+1) \times n$ matrix; this is necessary because the entries in the first row of $\tilde{\mathbf{X}}$ depend on noise variables with negative indices, up to and including Z_{1-n} . In order to formulate the transformation that maps \mathbf{Z} to $\tilde{\mathbf{X}}$ concisely in the next lemma, we define the matrices $K_n = \begin{pmatrix} 0 & 0 \\ \mathbf{1}_{n-1} & 0 \end{pmatrix} \in \mathbb{R}^{n \times n}$, as well as the polynomials $\chi_n(z) = c_0 + c_1 z + \dots + c_n z^n$ and $\bar{\chi}_n(z) = z^n \chi(1/z) = c_n + c_{n-1} z + \dots + c_0 z^n$.

Lemma 4. *With $\tilde{\mathbf{X}}$, \mathbf{Z} , K_n and $\chi_n, \bar{\chi}_n$ defined as before it holds that*

$$\tilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} \chi_n(K_n^T) \\ \bar{\chi}_n(K_n) \end{bmatrix}. \quad (10)$$

Proof. Let $s_N : \mathbb{R}^N \rightarrow \mathbb{R}^N$ be the right shift operator defined by $s_N(v_1, \dots, v_N) = (0, v_1, \dots, v_{N-1})$ and for positive integers r, s denote by $\text{vec}_{r,s} : \mathbb{R}^{r \times s} \rightarrow \mathbb{R}^{rs}$ the bijective linear operator that transforms a matrix into a vector by horizontally concatenating its subsequent rows, starting with the first one. The operator $S_{r,s} : \mathbb{R}^{r \times s} \rightarrow \mathbb{R}^{r \times s}$ is then defined as $S_{r,s} = \text{vec}_{r,s}^{-1} \circ s_{rs} \circ \text{vec}_{r,s}$. This operator shifts all entries of a matrix to the right except for the entries in the last column, which are shifted down and moved into the first column. For $k = 1, 2, \dots$, the operator $S_{r,s}^k$ is defined as the k -fold composition of $S_{r,s}$. In the following, we write $S := S_{p+1,n}$. With this notation it is clear that $\widetilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \chi_n(S) \mathbf{Z}$. In order to obtain Eq. (10), we observe that the action of S can be written in terms of matrix multiplications as $S\mathbf{Z} = K_{p+1}\mathbf{Z}E + \mathbf{Z}K_n^\top$, where the entries of the $n \times n$ matrix E are all zero except for a one in the lower left corner. Using the fact that $E(K_n^\top)^m E$ is zero for every non-negative integer m it follows by induction that $S^k, k = 1, \dots, n$, acts like

$$S^k \mathbf{Z} = \mathbf{Z}(K_n^\top)^k + K_{p+1} \mathbf{Z} \sum_{i=1}^k (K_n^\top)^{k-i} E (K_n^\top)^{i-1} = \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} (K_n^\top)^k \\ K_n^{n-k} \end{bmatrix}.$$

This implies that

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \sum_{k=0}^n c_k \begin{bmatrix} (K_n^\top)^k \\ K_n^{n-k} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} \chi_n(K_n^\top) \\ \bar{\chi}_n(K_n) \end{bmatrix}$$

and completes the proof. \square

While the last lemma gives an explicit description of the relation between \mathbf{Z} and $\widetilde{\mathbf{X}}$, it is impractical for directly determining the limiting spectral distribution of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$. The reason is that \mathbf{Z} appears twice in the central block-diagonal matrix and is moreover multiplied by some deterministic matrices from both the left and the right. The LSD of the product of three random matrices has been computed in the literature [25], but this result is not applicable in our situation due to the appearance of the random block matrix in Eq. (10). Sample covariance matrices derived from random block matrices have been considered in Rashidi Far et al. [19]. However, they only treat the Gaussian case and, more importantly, do not cover the case of a non-trivial population covariance matrix. We are thus not aware of any result allowing to derive the LSD of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ directly from Lemma 4.

The next proposition allows us to circumvent this problem. It is shown that, at least asymptotically and at the cost of slightly changing the size of the involved matrices, one can simplify the structure of $\widetilde{\mathbf{X}}$ so that \mathbf{Z} appears only once and is multiplied by a deterministic matrix only from the right.

Proposition 5. *Let \mathbf{Z} , K_n and $\chi_n, \bar{\chi}_n$ be as before and define the matrix $\widehat{\mathbf{X}} := \mathbf{Z}\Omega \in \mathbb{R}^{(p+1) \times (n+1)}$, where*

$$\Omega = \begin{bmatrix} 0 & \mathbf{1}_n & \mathbf{1}_n & 0 \end{bmatrix} \begin{bmatrix} \chi_{n+1}(K_{n+1}^\top) \\ \bar{\chi}_{n+1}(K_{n+1}) \end{bmatrix} \in \mathbb{R}^{n \times (n+1)}. \quad (11)$$

If the empirical spectral distribution of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$ converges to a limit, then the empirical spectral distribution of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ converges to the same limit.

Proof. In order to be able to compare the limiting spectral distributions of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ and $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$ in spite of their dimensions being different, we introduce the matrix $\overline{\mathbf{X}} = \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\mathbf{X}} \end{bmatrix} \in \mathbb{R}^{(p+1) \times (n+1)}$.

Clearly, $F^{p^{-1}\overline{\mathbf{X}}\overline{\mathbf{X}}^\top} = (p+1)^{-1}\delta_0 + p(p+1)^{-1}F^{p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top}$, which implies equality of the limiting spectral distributions provided either of the two, and hence both, exists. It is therefore sufficient to show that the LSD of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$ and $p^{-1}\overline{\mathbf{X}}\overline{\mathbf{X}}^\top$ are identical; this will be done by verifying the three conditions

of Lemma 2. The remainder of the proof will be divided in two parts. In the first part we check the validity of assumption i) about the difference $\widehat{\mathbf{X}} - \overline{\mathbf{X}}$, whereas in the second one we consider the terms $\text{tr} \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top$ and $\text{tr} \overline{\mathbf{X}} \overline{\mathbf{X}}^\top$, which appear in conditions ii) and iii).

Step 1. Using the definitions of $\widehat{\mathbf{X}}$ and $\overline{\mathbf{X}}$, it follows that

$$\begin{aligned} \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}} &:= \frac{1}{p^2} \text{tr} (\widehat{\mathbf{X}} - \overline{\mathbf{X}}) (\widehat{\mathbf{X}} - \overline{\mathbf{X}})^\top = \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} [\widehat{\mathbf{X}}_{ij} - \overline{\mathbf{X}}_{ij}]^2 \\ &\leq \frac{2}{p^2} \sum_{i=2}^{p+1} \sum_{j=2}^{n+1} \left[\sum_{k,k'=j}^n Z_{(i-2)n+k} Z_{(i-2)n+k'} c_{j-k+n+1} c_{j-k'+n+1} + \sum_{k,k'=j-1}^n Z_{(i-3)n+k} Z_{(i-3)n+k'} c_{j-k+n-1} c_{j-k'+n-1} \right] \\ &\quad + \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{k,k'=1}^n Z_{(i-2)n+k} Z_{(i-2)n+k'} c_{n-k+2} c_{n-k'+2} + \frac{2}{p^2} \sum_{j=2}^{n+1} \sum_{k,k'=1}^{j-1} Z_{-n+k} Z_{-n+k'} c_{j-k-1} c_{j-k'-1} \\ &\quad + \frac{2}{p^2} \sum_{j=2}^{n+1} \sum_{k,k'=j}^n Z_{-n+k} Z_{-n+k'} c_{j-k+n+1} c_{j-k'+n+1} =: \sum_{i=1}^5 \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(i)}, \end{aligned} \quad (12)$$

where the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ was used twice. In order to show that the variances of expression (12) are summable, we consider each term in turn. For the second moment of the first term of Eq. (12) we obtain

$$\begin{aligned} \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} &:= \mathbb{E} \left(\Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)} \right)^2 = \frac{4}{p^4} \sum_{i,i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k,k'=1}^{n-j+1} \sum_{m,m'=1}^{n-j'+1} \mathbb{E} [Z_{(i-1)n-k+1} Z_{(i-1)n-k'+1} Z_{(i'-1)n-m+1} Z_{(i'-1)n-m'+1}] \times \\ &\quad c_{j+k} c_{j+k'} c_{j'+m} c_{j'+m'}. \end{aligned}$$

As before we consider all configurations where above expectation is not zero. The expectation equals σ_4 if $i = i'$ and k, k', m, m' are equal, and, hence,

$$\mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} \leq \frac{4\sigma_4}{p^4} \sum_{i=2}^{p+1} \sum_{k=1}^n \left(\sum_{j=2}^{n+1} c_{j+k}^2 \right)^2 \leq \frac{4\sigma_4}{p^3} \sum_{k=1}^n \bar{c}_k^2 \left(\sum_{j=2}^{n+1} \bar{c}_j \right)^2 = O(n^{-3}).$$

The expectation is one if the four Z can be collected in two non-equal pairs. The first term equals the second, and the third equals the fourth if $k = k'$ and $m = m'$, and thus

$$\mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} = \frac{4}{p^4} \sum_{i,i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k=1}^{n-j+1} \sum_{m=1}^{n-j'+1} c_{j+k}^2 c_{j'+m}^2 - \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} = \frac{4}{p^2} \left(\sum_{j=2}^{n+1} \sum_{k=1}^{n-j+1} c_{j+k}^2 \right)^2 - \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} = O(n^{-2}).$$

Likewise, the contribution from pairing the first factor with the third, and the second with the fourth, can be estimated as

$$\left| \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} \right| \leq \frac{4}{p^4} \sum_{i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k,k'=1}^n |c_{j+k} c_{j+k'} c_{j'+k} c_{j'+k'}| + \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} \leq \frac{4}{p^3} \left(\sum_{j=1}^{n+1} \bar{c}_j \right)^4 + \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}} = O(n^{-3}).$$

Obviously, the configuration $\mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}}$ can be handled the same way as $\mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}}$ above. Thus we have shown that the second moment of $\Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(1)}$, the first term in Eq. (12), is of order n^{-2} . This can be shown

for the second term in Eq. (12) in the same way. We now consider the second moment of the third term in Eq. (12):

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}} := \mathbb{E} \left(\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)} \right)^2 = \frac{1}{p^4} \sum_{i,i'=1}^{p+1} \sum_{\substack{k,k'=1 \\ m,m'=1}}^n \mathbb{E} [Z_{(i-2)n+k} Z_{(i-2)n+k'} Z_{(i'-2)n+m} Z_{(i'-2)n+m'}] c_{n-k+2} c_{n-k'+2} c_{n-m+2} c_{n-m'+2}.$$

Distinguishing the same cases as before, we have $\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\prod \prod} = \sigma_4 \frac{p+1}{p^4} \sum_{k=1}^n c_{n-k+2}^4 = O(n^{-3})$ and, thus,

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\prod \prod} = \frac{(p+1)^2}{p^4} \left(\sum_{k=1}^n c_{n-k+2}^2 \right)^2 - \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\prod \prod} = O(n^{-2}),$$

as well as $\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\prod \prod} = \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\prod \prod} = O(n^{-3})$. Thus, the second moment of the third term in Eq. (12) is of order $O(n^{-2})$; repeating the foregoing arguments, it can be seen that the second moments of $\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(4)}$ and $\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(5)}$, the two last terms in Eq. (12), are of order $O(n^{-2})$ as well, so that we have shown that

$$\frac{1}{p^4} \mathbb{E} \left[\text{tr} (\widehat{\mathbf{X}} - \overline{\mathbf{X}}) (\widehat{\mathbf{X}} - \overline{\mathbf{X}})^\top \right]^2 = \mathbb{E} (\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}})^2 \leq 5 \sum_{i=1}^5 \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(i)}} = O(n^{-2}).$$

Step 2. In this step we shall prove that both $\Sigma_{\widehat{\mathbf{X}}} := p^{-2} \text{tr} \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top$ and $\Sigma_{\overline{\mathbf{X}}} := p^{-2} \text{tr} \overline{\mathbf{X}} \overline{\mathbf{X}}^\top$ have bounded first moments, and that their variances are summable sequences in n , i. e. we check conditions ii) and iii) of Lemma 2. Since $\text{tr} \overline{\mathbf{X}} \overline{\mathbf{X}}^\top$ is equal to $\text{tr} \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top$, the claim about $\Sigma_{\overline{\mathbf{X}}}$ has already been shown in the second step of the proof of Proposition 3. For the first term one finds, by the definition of $\widehat{\mathbf{X}}$, that

$$\begin{aligned} \Sigma_{\widehat{\mathbf{X}}} &= \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \left(\sum_{k=1}^{j-1} Z_{(i-2)n+k} c_{j-k-1} + \sum_{k=j}^n Z_{(i-2)n+k} c_{j-k+n+1} \right)^2 \\ &\leq \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=1}^{j-1} Z_{(i-2)n+k} c_{j-k-1} Z_{(i-2)n+k'} c_{j-k'-1} \\ &\quad + \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=j}^n Z_{(i-2)n+k} c_{j-k+n+1} Z_{(i-2)n+k'} c_{j-k'+n+1} =: \Sigma_{\widehat{\mathbf{X}}}^{(1)} + \Sigma_{\widehat{\mathbf{X}}}^{(2)}. \end{aligned}$$

Clearly, the first two moments of $\Sigma_{\widehat{\mathbf{X}}}^{(1)}$ are given by

$$\mu_{1,\Sigma_{\widehat{\mathbf{X}}}^{(1)}} := \mathbb{E} \Sigma_{\widehat{\mathbf{X}}}^{(1)} = \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=1}^{j-1} \mathbb{E} [Z_{(i-2)n+k} Z_{(i-2)n+k'}] c_{j-k-1} c_{j-k'-1} = \frac{2(p+1)}{p^2} \sum_{j=1}^{n+1} \sum_{k=1}^{j-1} c_{k-1}^2,$$

and

$$\begin{aligned} \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}} &:= \mathbb{E} \left(\Sigma_{\widehat{\mathbf{X}}}^{(1)} \right)^2 = \frac{4}{p^4} \sum_{i,i'=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k,k'=1}^{j-1} \sum_{m,m'=1}^{j'-1} \mathbb{E} (Z_{(i-2)n+k} Z_{(i-2)n+k'} Z_{(i'-2)n+m} Z_{(i'-2)n+m'}) \times \\ &\quad c_{j-k-1} c_{j-k'-1} c_{j'-m-1} c_{j'-m'-1}. \end{aligned}$$

We separately consider the case that all four factors are equal, and the three possible pairings of the four Z . If all four Z are equal, it must hold that $i = i', k = k' = m = m'$, with contribution

$$\begin{aligned}\mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} &= \frac{4\sigma_4}{p^4} \sum_{i=1}^{p+1} \sum_{j, j'=1}^{n+1} \sum_{k=1}^{\min\{j, j'\}-1} c_{j-k-1}^2 c_{j'-k-1}^2 \\ &\leq \frac{4\sigma_4(p+1)}{p^4} \sum_{j, j'=1}^{n+1} \bar{c}_{j-\min\{j, j'\}} \bar{c}_{j'-\min\{j, j'\}} \sum_{k=1}^{\min\{j, j'\}-1} \bar{c}_{k-1}^2 \leq \frac{4\sigma_4(p+1)}{p^4} \sum_{j, j'=1}^{n+1} \bar{c}_0 \bar{c}_{|j-j'|} \sum_{k=1}^n \bar{c}_{k-1}^2.\end{aligned}$$

Introducing the new summation variable $\delta_j := j - j'$, one finds that

$$\mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} \leq \frac{4\sigma_4(p+1)(n+1)}{p^4} \bar{c}_0 \left[\bar{c}_0 + 2 \sum_{\delta_j=1}^n \bar{c}_{\delta_j} \right] \sum_{k=1}^n \bar{c}_{k-1}^2 = O(n^{-2}). \quad (13)$$

The first factor being paired with the second, and the third with the fourth, means that $k = k', m = m'$, and $m \neq (i - i')n + k$, so that the contribution of this configuration is given by

$$\mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} = \frac{4}{p^4} \sum_{i, i'=1}^{p+1} \sum_{j, j'=1}^{n+1} \sum_{k=1}^{j-1} \sum_{m=1}^{j'-1} c_{j-k-1}^2 c_{j'-m-1}^2 - \mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} = \left(\mu_{1, \Sigma_{\bar{X}}^{(1)}} \right)^2 + O(n^{-2}). \quad (14)$$

For the $\square\square\square\square$ pairing, the constraints are $i = i', k = m, k' = m', k \neq k'$, and the corresponding contribution is

$$\begin{aligned}\mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} &= \frac{4}{p^4} \sum_{i=1}^{p+1} \sum_{j, j'=1}^{n+1} \sum_{k, k'=1}^{\min\{j, j'\}-1} c_{j-k-1} c_{j-k'-1} c_{j'-k-1} c_{j'-k'-1} - \mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} \\ &\leq \frac{4(p+1)}{p^4} \sum_{j, j'=1}^{n+1} \bar{c}_{j-\min\{j, j'\}} \bar{c}_{j'-\min\{j, j'\}} \sum_{k, k'=1}^{\min\{j, j'\}-1} \bar{c}_{k-1} c_{k'-1} + O(n^{-2}) \\ &\leq \frac{4(p+1)(n+1)}{p^4} \bar{c}_0 \left[\bar{c}_0 + 2 \sum_{\delta_j=1}^n \bar{c}_{\delta_j} \right] \sum_{k, k'=1}^n \bar{c}_{k-1} \bar{c}_{k'-1} + O(n^{-2}) = O(n^{-2}). \quad (15)\end{aligned}$$

Renaming the summation indices shows that $\mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square} = \mu_{2, \Sigma_{\bar{X}}^{(1)}}^{\square\square\square\square}$. Combining this with the displays (13) to (15), it follows that $\mathbb{V}\text{ar} \Sigma_{\bar{X}}^{(1)} = \mu_{2, \Sigma_{\bar{X}}^{(1)}} - \mu_{1, \Sigma_{\bar{X}}^{(1)}}^2 = O(n^{-2})$. Since a very similar reasoning can be applied to $\Sigma_{\bar{X}}^{(2)}$, and $p^{-4} \mathbb{V}\text{ar} \text{tr} \widehat{\mathbf{X}\mathbf{X}}^\top$ is smaller than $2 \mathbb{V}\text{ar} \Sigma_{\bar{X}}^{(1)} + 2 \mathbb{V}\text{ar} \Sigma_{\bar{X}}^{(2)}$, we conclude that $p^{-4} \mathbb{V}\text{ar} \text{tr} \widehat{\mathbf{X}\mathbf{X}}^\top$ is of order $O(n^{-2})$. \square

The intention behind Proposition 5 was to allow the application of results about the limiting spectral distribution of matrices of the form $\mathbf{Z}\mathbf{H}\mathbf{Z}^\top$, where \mathbf{Z} is an i. i. d. matrix, and \mathbf{H} is a positive semidefinite matrix. Expressions for the Stieltjes transform of the LSD of such matrices in terms of the LSD of \mathbf{H} have been obtained by Marchenko and Pastur [14], Silverstein and Bai [20], and, in the most general form, by Pan [17]. The next lemma shows that in the current context the population covariance matrix \mathbf{H} has the same LSD as the auto-covariance matrix Γ of the process X_t , which is defined in terms of the auto-covariance function $\gamma(h) = \sum_{j=0}^{\infty} c_j c_{j+|h|}$ by $\Gamma = (\gamma(i - j))_{i,j}$; this correspondence is used to characterize the LSD of \mathbf{H} by the spectral density f associated with the coefficients $(c_j)_j$.

Lemma 6. *Let Ω be given by Eq. (11). The limiting spectral distribution of the matrix $\Omega\Omega^\top$ exists and is the same as the limiting spectral distribution of the auto-covariance matrix Γ . It therefore satisfies*

$$\int h(\lambda) \hat{F}^{\Omega\Omega^\top}(\mathrm{d}\lambda) = \frac{1}{2\pi} \int_0^{2\pi} h(f(\omega)) \mathrm{d}\omega, \quad (16)$$

for every continuous function h .

Proof. The first claim follows by standard computations from the fact that Ω is, except for one missing row, a circulant matrix with entries $\Omega_{ij} = c_{n+j-i \bmod (n+1)}$, and Bai and Silverstein [5, Corollaries A.41 and A.42]. The second claim is an application of Szegő's limit theorem about the LSD of Toeplitz matrices; see Szegő [21, Theorem XVIII] for the original result or, e. g., Böttcher and Silbermann [7, Sections 5.4 and 5.5] for a modern treatment. \square

Proof of Theorem 1. According to Proposition 5, the matrix $\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$ is of the form $\mathbf{Z}\Omega\Omega^\top\mathbf{Z}^\top$, where Ω is given by Eq. (11). Using Pan [17, Theorem 1] and the fact that, by Lemma 6, the limiting spectral distribution of $\Omega\Omega^\top$ exists, it follows that the limiting spectral distribution $\hat{F}^{p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top}$ exists. Therefore, the combination of Propositions 3 and 5 shows that the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^\top$ also exists and is the same as that of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top$. Pan [17, equation (1.2)] thus implies that the Stieltjes transform of $\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^\top}$ is the unique mapping $s_{\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^\top}} : \mathbb{C}^+ \rightarrow \mathbb{C}^+$ which solves

$$\frac{1}{s_{\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^\top}}(z)} = -z + y \int_{\mathbb{R}} \frac{\lambda}{1 + \lambda s_{\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^\top}}(z)} \hat{F}^{\Omega\Omega^\top}(\mathrm{d}\lambda),$$

and Eq. (16) from Lemma 6 completes the proof. \square

4. SKETCH OF AN ALTERNATIVE PROOF OF THEOREM 1

In this section we indicate how Theorem 1 could be proved alternatively using the methods employed in Pfaffel and Schlemm [18]. We denote by $\widetilde{\mathbf{X}}_{(\alpha)}$ the matrix which is defined as in Eq. (1) but with the linear process being truncated at $\lfloor n^\alpha \rfloor$ with $0 < \alpha < 1$, i. e. $\widetilde{\mathbf{X}}_{(\alpha)} = \left(\sum_{j=0}^{\lfloor n^\alpha \rfloor} c_j Z_{(i-1)n+t-j} \right)_{it}$. If $1 - \alpha$ is sufficiently small, then an adaptation of the proof of Proposition 3 to this setting shows that $p^{-1}\mathbf{X}\mathbf{X}^\top$ and $p^{-1}\widetilde{\mathbf{X}}_{(\alpha)}\widetilde{\mathbf{X}}_{(\alpha)}^\top$ have the same limiting spectral distribution almost surely. The next step is to partition $\widetilde{\mathbf{X}}_{(\alpha)}$ into two blocks of dimensions $p \times \lfloor n^\alpha \rfloor$ and $p \times (n - \lfloor n^\alpha \rfloor)$, respectively. If we denote these two blocks by $\widetilde{\mathbf{X}}_{(\alpha)}^1$ and $\widetilde{\mathbf{X}}_{(\alpha)}^2$, i. e. $\widetilde{\mathbf{X}}_{(\alpha)} = [\widetilde{\mathbf{X}}_{(\alpha)}^1 \ \widetilde{\mathbf{X}}_{(\alpha)}^2]$, then clearly $\widetilde{\mathbf{X}}_{(\alpha)}\widetilde{\mathbf{X}}_{(\alpha)}^\top = \widetilde{\mathbf{X}}_{(\alpha)}^1 (\widetilde{\mathbf{X}}_{(\alpha)}^1)^\top + \widetilde{\mathbf{X}}_{(\alpha)}^2 (\widetilde{\mathbf{X}}_{(\alpha)}^2)^\top$, and an application of Bai and Silverstein [5, Theorem A.43] yields that

$$\sup_{\lambda \in \mathbb{R}_{\geq 0}} \left| F^{p^{-1}\mathbf{X}\mathbf{X}^\top}([0, \lambda]) - F^{p^{-1}\widetilde{\mathbf{X}}_{(\alpha)}^2 (\widetilde{\mathbf{X}}_{(\alpha)}^2)^\top}([0, \lambda]) \right| \leq \frac{1}{p} \text{rank} \left(\widetilde{\mathbf{X}}_{(\alpha)}^1 (\widetilde{\mathbf{X}}_{(\alpha)}^1)^\top \right) \leq \frac{1}{p} \min(\lfloor n^\alpha \rfloor, p) = O(p^{-1}n^\alpha) \rightarrow 0.$$

It therefore suffices to derive the limiting spectral distribution of $p^{-1}\widetilde{\mathbf{X}}_{(\alpha)}^2 (\widetilde{\mathbf{X}}_{(\alpha)}^2)^\top$. Since the matrix $\widetilde{\mathbf{X}}_{(\alpha)}^2$ has independent rows, this could be done by a careful adaptation of the arguments given in Pfaffel and Schlemm [18]. We chose, however, to provide a self-contained proof, which also provides intermediate results of independent interest like Proposition 5, and we therefore omit the lengthy details of this alternative proof.

REFERENCES

- [1] Anderson, G. W., Guionnet, A. and Zeitouni, O. [2010]. *An introduction to random matrices*, Vol. 118 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.

- [2] Anderson, G. W. and Zeitouni, O. [2008]. A law of large numbers for finite-range dependent random matrices, *Comm. Pure Appl. Math.* **61**(8): 1118–1154.
- [3] Aubrun, G. [2006]. Random points in the unit ball of l_p^n , *Positivity* **10**(4): 755–759.
- [4] Bai, Z. D. and Zhou, W. [2008]. Large sample covariance matrices without independence structures in columns, *Stat. Sinica* **18**(2): 425–442.
- [5] Bai, Z. and Silverstein, J. W. [2010]. *Spectral analysis of large dimensional random matrices*, Springer Series in Statistics, second edn, Springer, New York.
- [6] Bose, A., Subhra Hazra, R. and Saha, K. [2009]. Limiting spectral distribution of circulant type matrices with dependent inputs, *Electron. J. Probab.* **14**(86): 2463–2491.
- [7] Böttcher, A. and Silbermann, B. [1999]. *Introduction to large truncated Toeplitz matrices*, Universitext, Springer-Verlag, New York.
- [8] Bryc, W., Dembo, A. and Jiang, T. [2006]. Spectral measure of large random Hankel, Markov and Toeplitz matrices, *Ann. Probab.* **34**(1): 1–38.
- [9] Granger, C. W. J. and Joyeux, R. [1980]. An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal.* **1**(1): 15–29.
- [10] Hachem, W., Loubaton, P. and Najim, J. [2005]. The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity, *Markov Process. Related Fields* **11**(4): 629–648.
- [11] Hachem, W., Loubaton, P. and Najim, J. [2006]. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile, *Ann. Inst. H. Poincaré Probab. Statist.* **42**(6): 649–670.
- [12] Hofmann-Credner, K. and Stolz, M. [2008]. Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices, *Electron. Commun. Probab.* **13**: 401–414.
- [13] Hosking, J. R. M. [1981]. Fractional differencing, *Biometrika* **68**(1): 165–176.
- [14] Marchenko, V. A. and Pastur, L. A. [1967]. Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb. (N.S.)* **72**(114)(4): 507–536.
- [15] Meckes, M. W. [2007]. On the spectral norm of a random Toeplitz matrix, *Electron. Comm. Probab.* **12**: 315–325.
- [16] Mehta, M. L. [2004]. *Random matrices*, Pure and Applied Mathematics, third edn, Elsevier/Academic Press, Amsterdam.
- [17] Pan, G. [2010]. Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix, *J. Multivar. Anal.* **101**(6): 1330–1338.
- [18] Pfaffel, O. and Schlemm, E. Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.*, 31 (2):313–329, 2011.
- [19] Rashidi Far, R., Oraby, T., Bryc, W. and Speicher, R. [2008]. On slow-fading MIMO systems with nonseparable correlation, *IEEE Trans. Inform. Theory* **54**(2): 544–553.
- [20] Silverstein, J. W. and Bai, Z. D. [1995]. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices, *J. Multivar. Anal.* **54**(2): 175–192.
- [21] Szegő, G. [1920]. Beiträge zur Theorie der Toeplitzschen Formen, *Math. Z.* **6**(3-4): 167–202.
- [22] Wachter, K. W. [1978]. The strong limits of random matrix spectra for sample matrices of independent elements, *Ann. Probab.* **6**(1): 1–18.
- [23] Wigner, E. P. [1958]. On the distribution of the roots of certain symmetric matrices, *Ann. of Math. (2)* **67**(2): 325–327.
- [24] Yin, Y. Q. [1986]. Limiting spectral distribution for a class of random matrices, *J. Multivar. Anal.* **20**(1): 50–68.
- [25] Zhang, L. [2006]. *Spectral Analysis of large dimensional random matrices*, PhD thesis, National University of Singapore.

TUM INSTITUTE FOR ADVANCED STUDY & ZENTRUM MATHEMATIK, TECHNISCHE UNIVERSITÄT MÜNCHEN, BOLTZMANNSTRASSE 3,
85748 GARCHING BEI MÜNCHEN, GERMANY.